AD-A014 799

ASUPT (ADVANCED SIMULATION IN UNDERGRADUATE PILOT TRAINING) AUTOMATED OBJECTIVE PERFORMANCE MEASUREMENT SYSTEM

Wayne L. Waag, et al

Air Force Human Resources Laboratory
Brooks Air Force Base, Texas

March 1975

AFHRL-TR-75-3

AIR FORCE HUMAN RESOURCES LABORATORY

ASUPT AUTOMATED OBJECTIVE PERFORMANCE MEASUREMENT SYSTEM

By

Wayne L. Waag
Edward E. Eddowes
John H. Fuller, Jr., Capt, USAF
Robert R. Fuller, Major, USAF

FLYING TRAINING DIVISION
Williams Air Force Base, Arizona 85224

March 1975
Interim Report for Period July 1973 – June 1974

Approved for public release; distribution unlimited.

AIR FORCE SYSTEMS COMMAND
BROOKS AIR FORCE BASE, TEXAS 78235

NOTICE

When US Government drawings, specifications, or other data are used for any purpose other than a definitely related Government procurement operation, the Government thereby incurs no responsibility nor any obligation whatsoever, and the fact that the Government may have formulated, furnished, or in any way supplied the said drawings, specifications, or other data is not to be regarded by implication or otherwise, as in any manner licensing the holder or any other person or corporation, or conveying any rights or permission to manufacture, use, or sell any patented invention that may in any way be related thereto.

This interim report was submitted by Flying Training Division, Air Force Human Resources Laboratory, Williams Air Force Base, Arizona 85224, under project 1123, with Hq Air Force Human Resources Laboratory (AFSC), Brooks Air Force Base, Texas 78235.

This report has been reviewed and cleared for open publication and/or public release by the appropriate Office of Information (OI) in accordance with AFR 190-17 and DoDD 5230.9. There is no objection to unlimited distribution of this report to the public at large, or by DDC to the National Technical Information Service (NTIS).

This technical report has been reviewed and is approved.

WILLIAM V. HAGIN, Technical Director
Flying Training Division

Approved for publication.

HAROLD E. FISCHER, Colonel, USAF
Commander

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER AFHRL-TR-75-3	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) ASUPT AUTOMATED OBJECTIVE PERFORMANCE MEASUREMENT SYSTEM		5. TYPE OF REPORT & PERIOD COVERED Interim July 1973 – June 1974
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Wayne L. Waag, John H. Fuller, Jr., Edward E. Eddowes, Robert R. Fuller		8. CONTRACT OR GRANT NUMBER(s)
9. PERFORMING ORGANIZATION NAME AND ADDRESS Flying Training Division Air Force Human Resources Laboratory Williams Air Force Base, Arizona 85224		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS 62703F 11230108
11. CONTROLLING OFFICE NAME AND ADDRESS Hq Air Force Human Resources Laboratory (AFSC) Brooks Air Force Base, Texas 78235		12. REPORT DATE March 1975
		13. NUMBER OF PAGES ~~16~~ 15
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)		15. SECURITY CLASS. (of this report) Unclassified
		15a. DECLASSIFICATION DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

pilot performance measurement
proficiency measurement
pilot evaluation
objective measurement

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

To realize its full research potential a need exists for the development of an automated objective pilot performance evaluation system for use in the Advanced Simulation in Undergraduate Pilot Training (ASUPT) facility. The present report documents the approach taken for the development of performance measures and also presents data collected from two preliminary evaluation studies. The results indicated that the objectively derived measures: (1) correlate highly with instructor ratings, and (2) discriminate between pilots of different experience levels. These findings are encouraging and demonstrate the potential of the present approach for generating the needed automated objective pilot performance measurement system.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

SUMMARY

Problem

The Advanced Simulation in Undergraduate Pilot Training (ASUPT) facility is designed to be a research device capable of providing answers regarding the hardware design and effective use of flight simulators. Using state-of-the-art motion and visual systems, the relationship between simulator fidelity and training effectiveness, as well as the applicability of advanced training concepts are to be investigated. Since ASUPT was designed to be a research simulator, the development of an adequate performance measurement system becomes the foundation of the proposed program of research.

Approach

One of the salient characteristics of flying is that it is criterion-directed. The execution of any maneuver requires that a certain definable objective be met. The degree to which these objectives are met would appear to represent an adequate description of performance. In other words, a criterion-referenced approach to objective performance measurement is proposed. Consequently, for each maneuver, the criterion objectives must be defined in terms of parameters available within the simulator. Using this approach, a set of performance measures can be generated for each of the maneuvers to be flown in ASUPT.

Results

To evaluate the potential of the proposed approach, two preliminary studies were conducted. Measures were generated for seven basic instrument maneuvers of varying levels of difficulty. Pilots of different experience levels flew these maneuvers and were evaluated by experienced instructor pilots. The results indicated that, (1) instructor pilots were consistent in their subjective evaluations, (2) the objective measures correlated highly with the subjective evaluations, and (3) the objective measures discriminated between pilots of different experience levels.

Implications

The data suggest the approach taken to be a viable one. The basic assumptions of the measurement scheme were corroborated by the data. The results: (1) suggest instructor evaluations represent a useful criterion for developing objective measures, (2) indicate the objectively derived measures possess a high degree of validity, and (3) provide some insight into the manner in which instructors assign grades. Potentially fruitful areas of further research are discussed.

PREFACE

This document represents a portion of the research program on Project 1123, Flying Training Development, Dr William V. Hagin, Project Scientist; Task 112301, Development of Performance Measurement Techniques for Air Force Flying Training, Dr Wayne L. Waag, Task Scientist, being carried out by the Air Force Human Resources Laboratory, Flying Training Division, Williams Air Force Base, Arizona.

TABLE OF CONTENTS

LIST OF TABLES

ASUPT AUTOMATED OBJECTIVE PERFORMANCE MEASUREMENT SYSTEM

I. INTRODUCTION

The Advanced Simulation in Undergraduate Pilot Training (ASUPT) facility is designed to be a research device capable of providing answers regarding both the hardware design and effective use of flight simulators. Using state-of-the-art motion and visual systems, the relationship between simulator fidelity and the training effectiveness of these systems, as well as the applicability of advanced training concepts are to be investigated. Since ASUPT is designed to be a research simulator, the development of an adequate performance measurement system becomes an essential component of the research program. This report documents the approach to performance measurement system development which has been taken and presents the results of two brief validation studies.

Measurement System Requirements

The criterion for evaluating a flight simulation device is its training effectiveness. From past evidence indicating positive transfer effects, it is assumed that performance in the simulator will be positively related to performance in the aircraft. As a result, the thrust of the present effort is to develop measures which reflect performance in the simulator. In addition, it is possible to use pilot performance in the simulator as a criterion against which to investigate alternative simulator hardware configurations and training strategies.

One of the salient characteristics of flying is that it is criterion-directed. For the execution of any maneuver or sequence of maneuvers there are definable objectives which must be accomplished. The degree to which these objectives are met represents an adequate description of performance. In other words, a criterion-referenced approach to pilot performance forms the basis for the present effort. Following such an approach, it is apparent that the definition of criterion objectives is of foremost importance. Within the context of measurement development for ASUPT, the critical question is whether these criterion objectives can be stated in terms of parameters available within the simulator. In other words, can behavioral objectives be defined in terms of the state of the simulated aircraft and control inputs of the pilot?

Aside from the requirement to define performance in terms of observable behavioral objectives, there are other constraints which are applicable. The first is parsimony in the selection of simulator parameters to be sampled. Criterion objectives are to be defined using as few parameters as possible. Furthermore, they are to be sampled and analyzed on a real-time basis so that the resulting measurement and feedback are immediate. Since measurement will be an integral part of training students in ASUPT, it is also necessary that the resulting output be meaningful and easily interpreted by both instructors and students.

In summary, a criterion-referenced approach to measurement system development is to be pursued within the constraints of the following requirements:

1. Measures will assess the degree to which the criterion-objectives are met.
2. Measures will reflect only the most salient characteristics of performance.
3. Measures will be meaningful and interpretable to the user—the student and instructor pilot.
4. Measures will be generated on a real-time basis so that feedback is immediate.

Measurement Development for ASUPT

Although the criterion-referenced approach represents the rationale for present developmental efforts, a further set of assumptions have guided implementation on ASUPT.

1. *Measurement system development parallels skill acquisition in the student pilot.* Since student pilots acquire flying skills in a hierarchical fashion, measurement development should proceed in a similar manner. Simply stated, a building block approach is assumed in which measurements for basic flying skills are the first to be implemented.

2. *The focus of the measurement system is the individual flight maneuvers.* This level of measurement seems to be most consistent with present flying training syllabuses.

Preceding page blank

3. *Maneuvers can be conceptualized as integrated sequences of steady states and transitions.* The fundamental flight attitudes plus transitions from one attitude to another form the conceptual segments for most maneuvers (Meyer, Laveson, Weissman, & Eddowes, 1974). It is necessary to define behavioral objectives for each segment since they are likely to differ from one to another.

4. *Maneuver performance can be evaluated by two sets of parameters: those reflecting the state of the aircraft and those reflecting the control inputs of the pilot.* Superior performance is assumed to manifest three characteristics: (a) accomplishing the criterion objectives as defined by the aircraft state parameters, (b) avoiding excessive rates and acceleration forces so that the maneuver is executed smoothly, and (c) accomplishing these objectives with the least amount of effort; that is, by minimizing control inputs.

5. *The evaluation of performance for a given parameter involves a comparison of the obtained value with some ideal value.* For parameters reflecting aircraft state, the deviation from the ideal provides an index of error. Since the ideal is seldom attained, it is more realistic to define *acceptable* performance in terms of an empirically determined tolerance band about the ideal value. For parameters reflecting smoothness and control inputs, the adapted ideal represents a performance level characteristic of the highly experienced pilot.

6. *The implementation of the measurement system requires four phases of development:* (a) definition of criterion objective in terms of a candidate set of simulator parameters, (b) evaluation of the proposed set of measures for the purpose of validation and simplification, (c) specification of criterion performance by requiring experience instructor pilots to fly the maneuver in question, and (d) the collection of normative data using students as they progress through the training program.

Preliminary Evaluation – Study I

In keeping with the approach outlined previously, the first maneuver for which measures were developed was straight-and-level flight. Since the criterion objectives for this maneuver are well-defined–maintain constant altitude, airspeed and heading–it was felt that an evaluation of its measurement scheme would provide a good starting place to determine whether the approach held promise. Any performance assessment system, if it is valid, should have at least two characteristics. First, it should bear a positive relationship to expert opinion regarding the quality of performance. Second, it should reliably discriminate between subjects having differing levels of experience. The demonstration of such construct validity seemed necessary for the evaluation of the present measurement scheme.

Using the pre-programming capability in ASUPT, a simple test scenario was developed: (1) initialize the simulator to 15,000 feet, 160 knots, heading 090 degrees, (2) unfreeze the simulator and allow 10 seconds for the pilot to "settle down," (3) sample selected parameters for one minute, and (4) freeze the simulator. The parameters sampled were altitude, airspeed, heading, stick movement, throttle movement, elevator stick force, pitch rate, pitch acceleration, roll rate, roll accelerations, vertical velocity, and vertical acceleration. Mean deviation and root-mean-square (RMS) deviations were computed for altitude, airspeed, heading, and stick force. RMS scores were computed for the remaining parameters.

A seven-point rating form was developed for use by instructor pilots who were to provide qualitative evaluations. Five items were to be rated: (1) altitude control, (2) airspeed control, (3) heading control, (4) over-all level of performance, and (5) smoothness. For each evaluation, two raters were used–one inside the cockpit and another at the console who only observed the repeater instruments. In this manner, an estimate of inter-rater reliability could be computed.

Subjects for the initial evaluation were 12 Air Force employees (Flying Training Division, Air Force Human Resources Laboratory) and one student pilot. Piloting experience ranged from zero to 8,000+ hours. Each subject was given a two-minute practice/instruction period followed by five one-minute runs. Three instructor pilots were used as raters and were randomly alternated between the cockpit and console. All missions were flown using full motion and G-seat. Of the total 65 runs, one was lost due to system failure.

The means of the cockpit and console ratings, along with their correlations, are presented in Table 1. The high intercorrelations suggest the instructors to be highly consistent in their ratings. It is interesting to note that the lowest correlation is for smoothness. This is most likely due to availability of motion cues for the rater inside the cockpit. In any case, the high intercorrelations suggest the ratings to be a reliable criterion against which to validate the objective measures.

Table 1. **Comparison of Instructor Judgments at Cockpit and Console For Straight and Level**

Measure	Mean Cockpit	Mean Console	Inter-Rater Correlation
Altitude Control	4.7458	4.7031	.8722
Heading Control	4.7627	4.6562	.9257
Airspeed Control	4.5424	4.7344	.8535
Overall Rating	4.6271	4.6719	.9109
Smoothness	4.4574	4.6250	.6297

The next step was to determine whether the objectively derived measures would reliably predict the instructor ratings. The adopted criteria were the overall rating and smoothness rating. Table 2 presents the correlations between the objectively derived measures of performance and these instructor evaluations for both the cockpit and console raters. A glance at these results suggest several things. First, the correlations between the measures and instructor evaluations are fairly consistent for both the cockpit and console ratings. Second, and most importantly, substantial relationships exist between a number of the objectively derived measures and the instructor pilot (IP) subjective evaluations. Third, the measure RMS vertical velocity predicted both criteria quite well. These findings were highly encouraging—suggesting that these objectively derived measures do relate to the instructor's evaluation of performance.

Table 2. **Correlations of Objective Performance Measures With Instructor Evaluations for Straight and Level**

Measure	Overall Rating		Smoothness Rating	
	Cockpit	Console	Cockpit	Console
Mean Altitude Error	.6492	.7238	-.4645	-.4834
Mean Airspeed Error	.5826	.6515	.4365	.4331
Mean Heading Error	.5075	.5134	.4729	-.5253
RMS Altitude	.7803	-.8348	-.5750	-.6071
RMS Airspeed	.7690	.7739	-.5678	-.6193
RMS Heading	.6393	.6498	-.5931	-.6941
Mean Stick Force	.2740	.3859	.3717	.3560
RMS Stick Force	.4434	.3101	-.3249	-.2639
RMS Stick Movement	.0098*	.0237*	-.2789	-.3369
RMS Throttle Movement	.1643*	-.1855*	-.3320	-.3028
RMS Pitch Rate	.3168	-.2625	-.5019	-.5036
RMS Pitch Acceleration	.0781*	.1277*	-.2006*	-.1586*
RMS Roll Rate	.0371*	-.0500*	.0365*	.0931*
RMS Roll Acceleration	.1527*	.0976*	.0145*	.0009*
RMS Vertical Velocity	.7737	-.7560	-.7004	-.7277
RMS Vertical Acceleration	.4172	-.3770	-.4663	-.5194

* Nonsignificant.

Using a forward selection multiple regression procedure, subsets of variables were selected which were predictive of the criterion. An iterative procedure was used wherein variables were added to the prediction equation until the increment in explained variance became statistically nonsignificant. At this point, the variables in the prediction equation were eliminated from the predictor set and the procedure repeated. In this manner multiple sets of predictors were defined. The criteria adopted were the overall rating and smoothness rating obtained from within the cockpit. Using the equations developed against the cockpit ratings, an attempt was made to predict the ratings obtained at the console. The results of these analyses are presented in Tables 3 and 4.

Table 3. **Prediction of Overall Instructor Pilot Ratings For Straight and Level**

Measure		Cockpit	Console
	Set 1		
RMS Altitude		.9066	.9044
Mean Altitude Error			
Mean Stick Force			
RMS Throttle Movement			
	Set 2		
RMS Vertical Velocity		.8839	.8867
RMS Airspeed			
RMS Stick Movement			
RMS Pitch Rate			
	Set 3		
RMS Heading		.8303	.8178
Mean Airspeed Error			
RMS Stick Force			
RMS Pitch Acceleration			
RMS Vertical Acceleration			
	Set 4		
Mean Heading Error		.5075	.5134

Table 4. **Prediction of Smoothness Ratings For Straight and Level**

Measure		Cockpit	Console
	Set 1		
RMS Vertical Velocity		.7926	.8005
RMS Throttle Movement			
RMS Altitude			
RMS Pitch Rate			
	Set 2		
RMS Heading		.7155	.7086
Mean Altitude Error			
RMS Pitch Acceleration			
	Set 3		
RMS Airspeed		.6718	.7053
RMS Vertical Velocity			
	Set 4		
Mean Heading Error		.6248	.6496
Mean Airspeed Error			
RMS Stick Movement			
	Set 5		
Mean Stick Force		.5585	.5277
RMS Stick Force			

Four subsets of variables were selected which were predictive of the overall ratings. The first subset consisting of RMS altitude, mean altitude error, mean stick force, and RMS throttle movement yielded a multiple R of .9066. The correlation between the predicted score (using the equation developed for the cockpit rating) and the console rating was .9044. Similar degrees of correspondence were obtained for remaining subsets of predictions. For smoothness, five subsets of predictors were identified. Again, high degrees of correspondence were obtained between the multiple Rs developed for the cockpit ratings and the subsequent correlation between predicted scores and the console ratings.

The results of the first study were highly encouraging and seemed to warrant the following conclusions: First, instructor pilots are highly consistent in their evaluations of performance. Consequently, it is possible to use such evaluations as one criterion against which to validate objective measures of performance. Second, the objective measures of performance developed for straight and level will reliably predict instructor evaluations. The demonstration of such predictive validity suggested the approach taken to be a reasonable one. To further evaluate the proposed system, a second study was undertaken.

Preliminary Evaluation – Study II

Scenarios were developed and the performance measurement software written for the following maneuvers: change of airspeed; constant airspeed climbs/descents; rate climbs/descents; and the steep turn. Similar rating forms were developed for each maneuver and the simultaneous cockpit/console evaluation procedures followed. Four T-37 instructors were the raters–two alternating at the console and the other two alternating in the cockpit. Ten subjects were used in the second study, again representing a wide range of skills. Four student pilots in T-37 training, three T-37 IPs, and three civilians were included. Each subject flew the following set of maneuvers: three airspeed changes; one constant airspeed climb; one constant airspeed descent; one rate climb; one rate descent; and three steep turns. For each climb/descent a level-off to altitude was required.

Inter-rater correlations were computed for each maneuver and are presented in Table 5. As indicated, the data for climbs and descents were pooled. Overall inter-rater correlations were computed for categories which were rated for all maneuvers. The data indicates substantial agreement among the raters, especially for the overall and smoothness ratings, even though the values were somewhat less than obtained for straight-and-level. Several possible reasons for the lowered inter-rater correlations should be mentioned. First, the maneuvers in the second study were of increased complexity. Since these maneuvers require several transitions in addition to a steady state condition, the instructor's job of monitoring all the relevant parameters is increased. Likewise, the performance of transitions from one steady state to another increases the number of cues available to the rater within the cockpit. A second possibility concerns rater bias. It is possible that different subjective criteria were used in ratings of the T-37 IPs as opposed to students. An examination of the ratings of one of the IP's performance records yielded large discrepancies between the cockpit and console ratings. The objective measures appeared to agree with the cockpit ratings in that the performance was quite good. However, according to the console rater, the performance was considered to be unsatisfactory. Such data strongly suggest the possibility of rater bias.

Table 5. **Inter-Rater Correlations for Study II**

Measure	Change of Airspeed	CAS Climb/Descent	Rate Climb/Descent	Steep Turn	Overall
Altitude Control	.7714	.6334	.7740	.6935	.7279
Airspeed Control	.7763	.4818	.8078	.4238	.6758
Heading Control	.6121	.8987	.7534	+	*
Rate Control	+	+	.6272	+	*
Bank Control	+	+	+	.7743	*
Overall	.7741	.6822	.8045	.7721	.7716
Smoothness	.7311	.7055	.9661	.8592	.8368

+ Not computed for maneuver.
* Not computed since item not rated in all maneuvers.

Correlations between each of the objective measures and the overall and smoothness ratings for each maneuver were computed and presented in Table 6. In this case, the ratings from the IP in the cockpit were used as the criteria. A perusal of the data warrants several conclusions. As expected, parameters reflecting performance of the criterion objectives were most related to the ratings. Likewise, the RMS values about the ideal were correlated more highly than mean deviations. The only exception was RMS bank error. In this case, an error in the computing software was discovered, thereby invalidating the resulting measure.

A forward selection regression analysis was computed in an attempt to develop prediction equations for the overall steep turn rating. The steep turn was selected since it represented the most difficult of the maneuvers tested. The initial subset of seven variables selected by the procedure yielded a multiple R of .8820. This equation was then used to predict the console ratings. The obtained correlation between predicted and observed console ratings was .9137. Again, multiple subsets were isolated which were predictive of the criterion. Although not verified, it seems likely that similar sets of prediction equations could have been developed for the other maneuvers. In any case, it is certainly clear that the objective measures developed are highly predictive of instructor ratings, thereby demonstrating the validity of the present approach to measurement.

A further set of analyses were computed in order to relate the objective measures to the instructor evaluations. Performance records for all maneuvers were pooled and placed into groups according to the evaluation of the cockpit instructor. Four groups were defined according to the grade assigned of U, F, G, or E. Descriptive statistics were then computed for each of the four groups. The results are presented in Table 7. Measures for rate of climb and bank were deleted due to the small number of cases within each of the groups. An examination of the data warrants several conclusions. First, it is apparent that there are clearly defined trends for a number of the objective measures across the four groups. There are clear cut decreases in root mean square for airspeed, heading, and altitude as a function of the subjective ratings. Likewise, there are decreases in the variability across subjects. Consequently, lowered ratings are characterized by increasing within-subject error as well as increasing between-subject variability. Such data reflect the fact that there is one way to execute the maneuver correctly, but many ways to commit errors and therefore receive a lower evaluation.

The results also verified the assumption that superior maneuver performance involves the minimization of control inputs. Performances rated excellent were those which minimized the amount of stick movements and also minimized stick deviation from the null position. Furthermore, superior performance was also characterized by minimum amount of control force, or, in other words, the efficient use of trim. While stick inputs were found to be related to rated performance, throttle movements were not. Of the proposed measures of smoothness only pitch rate, vertical velocity and vertical acceleration were related to rated performance. Again, the pattern of decreasing means and variances was found.

As discussed previously, a requirement of the measurement system is that it reliably discriminates between pilots of different experience levels. In this case, depending on their previous flying experience, subjects were placed into one of two categories–high versus low experience–and descriptive statistics computed. The results are presented in Table 8. It should be pointed out that the statistics computed for each objective measure were based on data resulting from all the maneuvers in which that measure was computed. For example, the statistics for heading measures were based on all maneuvers except the steep turn. The results indicate that a substantial number of the objective measures will differentiate between the two groups. Generally, the inexperienced group was characterized by higher error scores and much greater variability. For the steady state parameters, the RMS error scores were more discriminative than the mean error scores.

Implications

The results of these two evaluation studies warrant a number of conclusions. First, the data suggest the approach taken to be a viable one. The assumptions concerning superior maneuver performance have been corroborated by the data. Experienced pilots, and likewise those performances rated excellent, were characterized by: (1) meeting the criterion objectives in terms of minimizing aircraft state errors, (2) minimizing rates and accelerations at least in the pitch and Z axes, and (3) minimizing control inputs. In other words, superior performance involves getting the job done, doing it smoothly, and with a least amount of effort.

Table 6. **Correlations of Objective Measures with Instructor Ratings for Study II**

Measure	Change of Airspeed Overall	Change of Airspeed Smoothness	CAS Climb/Descent Overall	CAS Climb/Descent Smoothness	Rate Climb/Descent Overall	Rate Climb/Descent Smoothness	Steep Turn Overall	Steep Turn Smoothness
Mean Airspeed Error	–.1658	–.0636	–.0164	–.0470	–.0681	.0545	–.2322	–.1776
RMS Airspeed	–.5854	–.6856	–.4104	–.0702	–.4776	–.4951	–.6077	–.6112
Mean Heading Error	.2804	.1377	.0527	.0594	–.6670	–.7620	+	+
RMS Heading	–.5463	–.4144	–.5090	–.3614	–.6649	–.7899	+	+
Mean Altitude Error	–.6747	–.6386	.2130	.3374	–.0946	.0108	.5580	.5400
RMS Altitude	–.7085	–.6512	–.5690	–.4755	–.6868	–.7193	–.5435	–.5320
Mean Bank Error	+	+	+	+	+	+	–.5375	–.6121
RMS Bank	+	+	+	+	+	+	–.0384*	–.2136*
Mean Rate Error	+	+	+	+	–.3194	–.3530	+	+
RMS Rate	+	+	+	+	–.8029	–.8312	+	+
Mean Stick Force	–.1542	.0959	.1021	.0799	–.0199	.0013	.1282	.1761
RMS Stick Force	–.2898	–.1401	–.2475	–.1343	–.4920	–.6303	–.3475	–.4145
RMS Stick Movement	–.0047	–.2823	–.2416	–.2901	–.6836	–.7036	–.4787	–.6483
RMS Throttle Movement	.0463	–.0274	.3325	.1081	–.1729	–.3332	.1395	–.1279
RMS Fore-Aft Stick Position	–.2664	–.2394	–.1641	–.2295	–.2851	–.4081	–.3122	–.3969
RMS Lateral Stick Position	–.2467	–.4875	–.6364	–.6090	–.6630	.4763	–.6380	–.7771
RMS Pitch Rate	–.1145	–.3372	–.1161	–.1134	–.6315	–.8145	–.5375	–.6121
RMS Pitch Acceleration	.1988	.0411	.0753	.0839	–.5514	–.7437	–.0385	–.2137
RMS Roll Rate	–.0203	–.1202	–.0490	.0016	.0406	.0709	.0506	.0723
RMS Roll Acceleration	-.2311	–.2214	.3484	.3378	–.1827	–.0618	.0953	.0857
RMS Vertical Velocity	–.6069	–.6120	–.2559	–.4762	–.5170	–.5205	–.4225	–.4521
RMS Vertical Acceleration	–.3130	–.5645	–.3286	–.2582	–.7136	–.6975	–.4528	–.5116

* Measure incorrectly computed due to software error.
\+ Not computed for maneuver.

Table 7. **Descriptive Statistics of Objective Measures for Each Rating Category**

Measure	Means				Standard Deviations			
	U	F	G	E	U	F	G	E
Mean Airspeed	6.2029	2.6395	.8516	.9805	8.0183	5.3732	3.4874	1.5545
RMS Airspeed	12.3590	6.4086	3.9006	2.2694	7.3095	3.6592	2.0820	1.3643
Mean Heading	−1.8377	2.1337	−.9720	−.0938	7.2208	5.2646	3.0924	1.4064
RMS Heading	8.0273	5.0833	3.2715	1.4535	4.5197	3.9316	1.7367	.7454
Mean Altitude	−291.4250	−46.1848	−21.2107	4.9002	718.3638	125.3819	59.7469	19.9138
RMS Altitude	686.9624	133.9150	63.1433	27.5617	577.0505	82.5278	40.6938	14.7862
Mean Stick Force	−2.2373	−1.2741	−.6847	−.7125	3.4299	1.7442	1.9502	1.2228
RMS Stick Force	5.2705	2.8326	2.4151	1.6180	3.6954	1.8237	1.6616	1.2253
RMS Stick Movement	.3910	.2610	.1955	.1833	.1835	.1507	.0970	.0841
RMS Throttle Movement	.6958	.6254	.7671	.7899	.2190	.2884	.4653	.6258
RMS Fore-Aft Stick Position	3.2308	2.2098	1.9672	1.7736	1.8311	.8447	.6526	.5203
RMS Lateral Stick Position	.6336	.3687	.2848	.1903	.2870	.1940	.1372	.0964
RMS Pitch Rate	4.9255	2.0146	1.1146	1.1477	4.3768	1.6320	.8628	1.0557
RMS Pitch Acceleration	29.7499	20.9420	17.7711	20.1213	19.5700	19.8789	23.4170	19.4427
RMS Roll Rate	.4012	5314	.5801	.5197	.3163	.2967	.3373	.3062
RMS Roll Acceleration	.5806	.5286	.6199	.4462	.2333	.3462	.3637	.3160
RMS Vertical Velocity	1205.0500	337.6677	123.8052	168.8898	1789.3610	222.8050	184.1357	214.3813
RMS Vertical Acceleration	3.2379	1.5558	1.4965	1.2012	2.0841	.8067	.7379	.5517

Table 8. **Comparison of Objective Measures for Experienced And Inexperienced Pilots**

Measure	Experience		Inexperienced	
	Mean	S.D.	Mean	S.D.
Mean Airspeed	1.4730	2.5232	2.3486	6.1876
RMS Airspeed	3.4929	2.0878	6.7856	5.0718
Mean Heading	−.1843	2.2824	.6084	5.4465
RMS Heading	2.1827	1.4071	5.1275	3.6841
Mean Altitude	−21.6591	77.8214	−71.5790	316.2510
RMS Altitude	67.5192	67.3319	193.8939	322.1645
Mean Climb Rate	.1610	1.4210	.9480	2.7849
RMS Climb Rate	2.3482	.8100	4.4623	1.5058
Mean Bank	1.4191	.8855	4.1372	3.5044
Mean Stick Force	−.9480	1.9727	−1.0806	2.0091
RMS Stick Force	2.4354	1.7274	2.8486	2.3823
RMS Stick Movement	.1847	.1009	.2778	.1514
RMS Throttle Movement	.6668	.4829	.7679	.3872
RMS Fore-Aft Stick Position	2.0347	.6661	2.1971	1.1259
RMS Lateral Stick Position	.2517	.1356	.3968	.2294
Pitch Rate	1.0769	.9846	2.3859	2.4821
Pitch Acceleration	18.2833	23.9736	22.1778	18.4969
Roll Rate	.5146	.3078	.5749	.3257
Roll Acceleration	.5670	.3494	.5420	.3432
Vertical Velocity	172.3421	208.9697	409.9243	825.0552
Vertical Acceleration	1.5010	.7557	1.7047	1.2598

Second, the data indicate that instructor evaluations are a usable criterion for future measurement system development. The relatively high levels of agreement between the cockpit and console ratings are particularly encouraging since the availability of cues was radically different for each rater. The console rater only had access to the repeater instruments while the instructor in the cockpit could observe the students' behavior in addition to the flight instruments. Furthermore, kinesthetic cues were also available to the cockpit rater. The importance of these different cue sources in instructor evaluations is an area which should be addressed in future research studies.

Third, the objectively derived measures were shown to possess a certain degree of validity. Significant correlations between these measures and instructor evaluations were obtained. Furthermore, the measures were shown to discriminate between pilots of different experience levels. These results are most encouraging and convincingly demonstrate the fruitfulness of the present approach.

Fourth, the results provide some insight into the manner in which instructors assign grades. A hierarchical model seems most consistent with the data. To receive an excellent (E) rating, errors on the critical parameters must all be low. As the quality of the rating decreases, the potential for different errors increase. In fact, it is possible to commit error involving one parameter, control the others quite well, and still receive a low evaluation. This suggests that both the number and degree of error are important. The investigation of instructors grading strategies is another prime area for future research.

Fifth, the results provide preliminary data concerning the simplification of the present set of parameters. As expected, parameters reflecting performance of the criterion objectives yielded the highest validity coefficients. However, measures of mean error were not as effective as root-mean-square error. Furthermore, a number of the proposed measures of smoothness did not produce any significant relationships. Roll rate, roll acceleration, and pitch acceleration were not related to either instructor ratings or experience level. Likewise, throttle movements were not found to be important.

Taken as a whole, the results of these preliminary investigations are encouraging. The demonstrated validity of measures, developed for these basic instrument maneuvers, indicate the fruitfulness of the present approach. Efforts are currently underway for the development of performance measures for more complex maneuvers. It is expected that the resulting measurement system will meet both the research and training needs for future studies to be accomplished in ASUPT.

REFERENCES

Meyer, R.P., Laveson, J.I., Weissman, N.S., & Eddowes, E.E. *Behavioral taxonomy of undergraduate pilot training tasks and skills: Executive summary.* AFHRL-TR-74-33(I). Williams AFB, Ariz.: Flying Training Division, Air Force Human Resources Laboratory, December 1974.